Dylan Mumms | Lauren Setzer | Dalvin Parks

Goal

To predict the favorable attributes of a song, using Spotify and Billboard Hot 100 data, that contributes to the popularity and performance among listeners
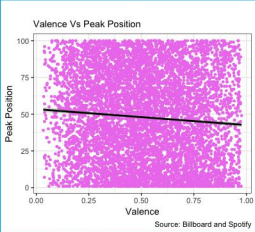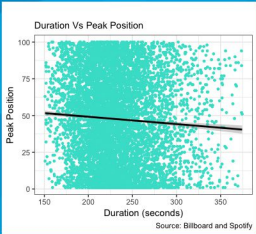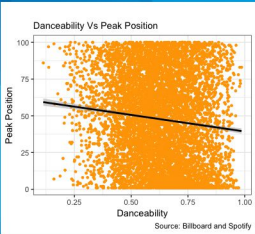
The Process

A combination of Spotify's song characteristics (valence, duration, danceability, instrumentalness, etc.) & Billboard Hot 100's rank & year of release were selected as the predictors & response variables respectively. The raw CSVs were loaded & converted into a standalone dataframe consisting of strictly numeric data in reference to completed characteristics

Results

The predictors & response's relationship was primarily depicted using a linear graph model, utilizing a cross validation lasso approach to hone in on the most significant features. A 10-fold cross validation approach resulted in an average adjusted mean squared error of 887.6.

Data

Danceability Vs Peak Position

Source: Billboard and Spotify

Duration Vs Peak Position

Source: Billboard and Spotify

Valence Vs Peak Position

Source: Billboard and Spotify

```
Deviance Residuals:
    Min      1Q   Median      3Q     Max
-57.957  -26.480  -0.654  25.468  65.036

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -8.981e+02  1.444e+02  -6.218 5.32e-10 ***
danceability -2.035e+01  2.607e+00  -7.807 6.70e-15 ***
year         4.833e-01  7.164e-02   6.746 1.64e-11 ***
duration_ms -4.222e-05  8.229e-06  -5.131 2.96e-07 ***
valence     -5.148e+00  1.782e+00  -2.888  0.00389 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

What makes a song popular?